STT-TTS Conversion Engine for Facilitating impaired people to Communicate on normal voice call

S.Suresh
Assistant Professor,
Department of CSE,
Panimalar Engineering College,Chennai,India.
m.suresh.suresh@gmail.com

Rino Simon T M, Max Sherwin, Jagadeeshwar Naidu G
Under Graduate students,
Department of CSE,
Panimalar Engineering College,Chennai,India.
rinosimon@gmail.com, maxsherwinn@gmail.com, jagadeesh2113@gmail.com

Received 18 February 2018 Received in revised form 21 February 2018 Accepted 22 February 2018
Available online 02 March 2018

ABSTRACT

The disabled (hearing and speech impaired people) find it difficult to communicate through a normal voice call. They require the ability to talk/listen remotely and this is unavoidable in emergency situations. To rectify this problem we propose a speech recognizing and conversion android application that has two major modules- a speech to text engine and a text to speech engine. The former receives an incoming call as the input and converts it to text which will be displayed on an interface, the latter receives the text on the interface and uses android speech synthesis to generate a voice and transmits it to the other end. It is to be noted that it's not mandatory for both users in the communication to have the app, only the disabled people will have the app.

Key words – Android Platform, Java Programming Language, Speech to Text, Text to Speech, Voice recognition, Hidden Markov models.

I. INTRODUCTION

Living with a disability is not easy, but most people who have some form of disability generally develop other heightened senses and skills, and are able to live an almost normal life and contribute to society. But in emergency situations they find it difficult to communicate with a remote person on a normal voice call. Mostly the hearing and speech impaired people face this difficulty. We propose an android application that can provide a solution. It uses the voice of the incoming call for a speech to text conversion .With modern processes, algorithms, and methods we can process speech signals easily and use it in our desirable fields. Our speech-to-text engine directly converts speech to text. Text-to-speech convention transforms linguistic information stored as data or text into speech. It is widely used in audio reading devices for blind people now a days [1].In the last few years however; the use of text-to-speech conversion technology has rapidly grown and is used for digital voice storage for voice mail and voice response systems It can also play a defining role in establishing communication of the speech impaired if it is incorporated into mobile phones so that text messages typed on our application interface could be converted into speech.

II.METHODOLOGY

To build the application we need to implement the following modules

- A Speech to Text module
- A text to speech module
- Creating an API

III. SPEECH TO TEXT MODULE

With modern processes, algorithms, and methods we can process speech signals easily and recognize the text. In this system, we are going to develop an on-line speech-to- text engine. The system acquires speech at run time through a microphone and processes the sampled speech to identify the uttered text.

Speech recognition system:

The speech recognition algorithm synthesises the voice received through the phone call and converts it to text. The recognized text can be stored in a file. Speech recognition is done via the Internet, connecting to Google's server.

It is divided into several blocks:

- Feature extraction.

- Acoustic models database based on the
- Training data.
- Dictionary.
- Language model.

The application is adapted to input messages in English. Speech recognition for Voice uses a technique based on Hidden Markov models (HMM – Hidden Markov Model). It is currently the most successful and most flexible approach to speech recognition. HMM algorithm is briefly described in this part. Process involves the conversion of acoustic speech into a set of words and is performed by software component. Speech recognition system can be divided into several blocks: feature extraction, acoustic models database which is built based on the training data, dictionary, language model and the speech recognition algorithm. Feature vectors from training database are used to estimate the parameters of acoustic models. Acoustic model describes properties of the basic elements that can be recognized. The basic element can be a phoneme for continuous speech or word for isolated words recognition. Dictionary is used to connect acoustic models with vocabulary words. Language model reduces the number of acceptable word combinations based on the rules of language and statistical information from different texts.

IV. TEXT TO SPEECH

The presented research aims at developing a working model of speech synthesizer for English in android based mobile phones along with creation of a light weight English speech database for android mobiles. The work will create a user-friendly environment to present the application effectively. The major requirement about implementing the work is we need a library of English text to its phoneme equivalent. There are number of such libraries available online. We can get these libraries by performing an online search. The TTS conversion is implemented for the mobile android environment. It is under the NLP and provides easy communication for the person who cannot speak but can communicate verbally by using this application. The crucial focus of this work is to develop a working model of speech synthesis for English script for android based mobile phones. Secondly, to create a light weight English speech database for android mobiles. We define the present work for English as well as for the regional language. TTS is the artificial production of human speech. It converts normal language text into speech. A TTS engine converts text in the written form to a phonemic representation and then it converts the phonemic representation to waveforms that can be output as sound. Front end and back end are the two parts of a TTS engine. NLP is an area of research and application that explores how computers can be used to understand and manipulate natural language text or speech. The module of general Text To Speech (TTS) conversion system consists of pre-processor, text analyser, morphological analyser, contextual analyser, syntactic prosodic parser, letter to sound module and prosody generator. A text analyser block is composed of a pre-processing module, which organizes the input sentences into manageable lists of words. It identifies numbers, abbreviations, idiomatic and transforming them into full text when needed. A morphological module performs task to propose all possible part of speech categories for each word taken individually, on the basis of their spelling. Inflected, derived and compound words are decomposed into their elementary graphemes units by simple regular grammars exploiting lexicons of stems and affixes. The contextual analyser module considers words in their context, which allows it to reduce the list of their possible part of speech of neighbouring words. Finally, a syntactic parser which examines the remaining search space finds the text structure that is more closely related to its expected prosodic realization.

V. CREATING THE INTERFACE

An API provides a set of functions and procedures that allows the creation of applications which access the features or data of an operating system, application, or other service. In our proposed system an API is required to provide our application with accessibility to the incoming voice call and transfer it to the Speech to Text engine. If the client responds through text the TTS engine converts the text to computer generated voice, now the API is responsible for mapping this voice as the output response for transmission. In case the client can speak he would reply through voice as in a normal voice call and there is no need for speech conversion. Thus, the API should also be able to recognize when the source for the output stream must be changed and react accordingly.

VI. TECHNICAL SPECIFICATION

The application to be developed is designed using the 'Eclipse' software. The reason for using Eclipse is that it is user-friendly and easy to design. The application, text-to- speech and speech-to- text (for the speech impaired and hearing impaired people) is to be free of cost. The application is to be developed using Java programming language. The user interface (UI) is to be designed using xml files. The application is expected to function in android devices, with API level 10 and above.

REFERENCES

1.Chen S, Song Q. Perimeter-based defense against high bandwidth DDoS attacks, IEEE Transactions On Parallel and Distributed Systems. 2005 Jun; 16(6):526–37.

2. Achar RK, Babu MS, Arun M. Border gateway protocol performance and its protection against disturbed denial of service attack. Indian Journal of Science and Technology. 2015 Jan; 8(S2). DOI: 10.17485/ijst/2015/v8iS2/59169.

3. Jin G, Yang J. Deterministic packet marking based on redundant decomposition for IP traceback. IEEE Communications Letters. 2006 Mar; 10(3):204–6.

4. Uddin M, Alsaqour R, Abdelhaq M. Intrusion detection system to detect DDoS attack in Gnutella hybrid P2P network. Indian Journal of Science and Technology. 2013 Feb; 6(2). DOI: 10.17485/ijst/2013/v6i2/30585.

5. Gong C, Sarac K. A more practical approach for single packet ip traceback using packet logging and marking. IEEE Transactions on Parallel and Distributed Systems. 2008 Oct; 19(10):1310–24.